DOCUMENT RESUME

ED 404 985 IR 018 243

AUTHOR Nigohosian, Robert H.
TITLE Scholarly Internet Research: Is it Real?
PUB DATE 3 Mar 96
NOTE 11p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Electronic Mail; *Evaluation; Higher Education; Information Literacy; *Information Seeking; Information Skills; *Information Sources; *Internet; *Online Searching; Research Needs; Student Research; User Needs (Information); World Wide Web
IDENTIFIERS Browsing; Information Overload

ABSTRACT

This paper addresses several issues relative to Internet research and ethics in an era of information explosion and information anxiety. Situations are reviewed in which misleading or "junk" information appears in web sites--information which novice users may not be able to distinguish from valuable sources. Suggestions for proper evaluation of Internet information are then outlined. The discussion includes: (1) awareness of electronic mail hoaxes and home pages that may be nothing more than a form of vanity publishing (it is noted that a document that may have information of great value, but one which has not been through the peer review process or disseminated by the trade publishing industry is still a vanity work); (2) identification of research needs and the usefulness of information sources; (3) validation of authors by investigating information such as reputation, perspective, timeliness, and costs for material; and (4) advantages of valid Internet research, or the ability for professors to spot check sources while sitting at their computer. Some search engines, which can be dependable vehicles for verification and evaluation of sources, are described: W3 Virtual Libraries Project; Clearinghouse Project; Yahoo; Point Communications; the McKinley Internet Directory; and Gale Guide. A review of some examples of search tool evaluation is also provided. (AEF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Scholarly Internet Research: Is It Real?

Robert H. Nigohosian
http://www.slcc.edu/b10/~~resume.htm~~ eca 274 / resume-h.htm

Department of Finance and Economics
Salt Lake Community College

ED 404 985

U.S. DEPARTMENT OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Robert H. Nigohosian
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

The Internet has evolved into a powerful tool for research, especially over the last several years. But in some ways it is a funny place that does not fit into a customary box of procedures and tools well known to information professionals of yesteryear. This network of networks is viewed by some as an unstructured information resource, with no overall control, a mix of services, no standard format, no subject organization or comprehensive index similar to printed and electronic sources already familiar to librarians. Although these are subject indexes, their multiplicity and fundamental differences make use confusing at times, raising the questions: Which one is best? Do we need to use them at all? What exactly are these indexes searching? Is there any one group who is responsible for their coordination and validation? (Ian R. Winship, "World Wide Web Searching Tools --An Evaluation"; World Wide Web address http://www.bubl.bath.ac.uk/BUBL/IWinship.html).

PURPOSE

This paper seeks to address several issues relative to Internet research and its ethical formation through this new era of information explosion, and resulting information anxiety. First, it may be interesting to review some situations in which misleading information may appear. Secondly, the question about responsiblity for management of evaluation and validation is discussed. Finally, a review of some examples of search tool evaluation is provided, in light of their usefulness in providing ethical and proper education for our students.

For the last year, with the upgrades of accessibility to Netscape, I have encouraged my students to use the World Wide Web (WWW) as a research tool in gathering information for papers they present in our course, Economic History of the United States. I was amazed at the availability of sources for various subjects the students selected, but one daunting feature continued to surface as we went from search to search. That feature was, and is growing in record numbers daily, the presence of "junk" information.

For example, one student went to Webcrawler and typed "slavery" into the search box. It was fascinating to see that of the 25 "hits" that were returned to the viewer, only about 3 were of any value at all. More interesting was the listing of "S & M Shops in Seattle" !!!!!! I guess slavery occurs there also!

IR018243

Do we need to discuss what the students in my Macroeconomics class found when they did a search on "bonds"? Obviously, some instruction was required to get them to specifiy the types of bonds, or else they would have wound up in Seattle again!

The evidence is clear: the Internet has enabled a whole new group to enter the world of publishing - those who did not learn the culture of the print publishing trade. Does anyone have the responsibility to explain the rules to new publishers, just as the Internet community inculcates new users with the Internet etiquette rules of the road?

For example, consider a web site devoted to Gilbert & Sullivan. Hope Tillman, in her presentation at the John F. Kennedy School of Government, Harvard University, made some observations:

There is a pretty clear table of contents. The welcome message from the Introduction points to the Savoynet listserv as well as to short bios of Wiliam S. Gilbert and Arthur S. Sullivan.

"Welcome to the Gilbert and Sullivan Archive. This archive is devoted to the works of William S. Gilbert and Arthur S. Sullivan, and is operated as a service to Gilbert & Sullivan fans by members of SavoyNet distribution list. The G & S Archive was established in September, 1993, by several SavoyNet members. It includes a variety of G & S related items, including clip art, librettos, song scores, and newsletter articles. New items are being added regularly."

An information professional might ask: Is there a reason that Boise State University should host this web site, inaddition to MIT or Harvard-Radcliffe, that also have web site? Preliminary research showed that there is a music department at Boise State, but it does not host a Gilbert & Sullivan festival or even performances listed during the current year at Boise State, according to their calendar of musical performances upcoming. It just so happens that a G & S afficianado is an associate professor in the Math Department and has obviously been instrumental in the hosting of the site there. What is the authority of the moderators? Jim Farron and Alex Feldman are listed. Only Alex Feldman is at Boise State, listed as an Associate Professor in the Department of Mathematics and Computer Science, with professional interests including Theory of Computation and Recursion Theory. From the web site, the identity of Jim Farron could not be determined. (Hope Tillman, "Finding Quality on the Internet or a Needle in a Haystack?"; prepared for presentation at the NEASIS program, "Evaluating the Quality of Information on the Internet" at the John F. Kennedy School of Government, Harvard University, Cambridge, Massachusetts, September 6, 1995).

Apart from the comedic and apparently insignificant connection some of these sites seem to have to our subjects of research, the requirement of proper validation and evaluation of sources became more and more apparent to our classes. This factor appears to be growing daily, as numerous homepages and commercial sites proliferate the landscape of the Web. Novice users are fair game for tricksters and tactless entrepreneurs that disguise their pages as valid sources of information when they may sometimes be nothing more than conjecture, opinion, and manipulated statistical reporting.

However, for its vagaries and flaws, the Internet still hosts an incredible amount of useful and

scholastic information. Has it become the instructor's responsibility to educate student researchers regarding proper evaluation and validation tools which can be used in guiding them to scholarly and verifiable information, and away from useless and confusing "junk"? Consider the following sugestions for proper research evaluation of Internet information:

1) Some "home page" publishing may be nothing more that a form of vanity publishing. (Hope Tillman, p. 1, " Finding Quality on the Internet, or a Needle in a Haystack"?) This may even include sites where an individual decides to share working papers or information they have been working on for a dissertation. Some "home pages" may appear as scholarly journal type articles, but may be disguised and manipulated information published by an individual posing as a professor from an institution of higher education. On the other hand, many home pages have been through a rigorous review process and should not be equated with the term "vanity".

What is a "vanity" work? This may be a very specific document that has information of great value but it hasn't been throught the peer review process intrinsic to scholarship or it hasn't been disseminated by the trade publishing industry. Prior to the information explosion promulgated by the accessibility of Netscape and Internet, vanity and short-run publishing has been possible in print and can be "quality" in nature, although that may not be as easy to determine without analysis. (Ibid.) Depending upon the curriculum, some instructors limit student researchers to manufacturers "home pages", for example. In a UNIX class taught by Bruce Worthen of Salt Lake Community College, students are not allowed to use any other home pages, and are cautioned to review http addresses that have a "~" in their address, as these are most likely to be found in personal home pages. Other borderline sources to be aware of may be those sources displaying an address containing "xmission.com" or "compuserve.com", or "aol.com". These sources may indeed be validated and scholarly, but students are advised to proceed with caution when confronting such addresses.

Janet Hovorka and Keith Slade, in their web site paper entitled "Evaluating and Citing Sources: Internet Truth or Fiction" (http://www.slcc.edu/lr/library/intwork.htm) cite several common E-Mail Hoaxes that pervade the Internet. For example, *Microsoft Bought the Catholic Church*. There were people that actually believed this one. Some stories stated that Microsoft was buying the church outright, while others stated that only the art collection had been secured. Or, *The Good Times Virus*. According to this hoax, if you open an e-mail message with the subject line: Good Times you will get this virus. However, e-mail comes in ascii which can not transmit a virus. The methods for discovering such hoaxes include: 1) checking the date - April 1 is probably a hoax. 2) Check the sender: for messages posted on a Usenet group, check the address of the sender at the top of the screen. Message posters can also be checked our on Deja News at http://www.dejanews.com. This site will tell the user what messages have been posted to what newsgroups by the sender.

2) Identification of Research Needs:

Students may streamline their searches by using four major steps in the research process:

a. Clearly identify the research need
b. Identify which types of resources you hope to find
c. Identify and use the search tools to find the information most appropriate for your need
d. Carefully and critically evaluate the information you have found

If, for example, students cannot identify the usefulness of an information source immediately, it should be considered a low priority to save, print or read online. Ask the question:
What kind of information is it?
Is it facts or opinions?
Is there any documentation, such as bibliography, footnotes, credits, or quotations? (C. Hansen, "Internet Navigator - Resource Discovery; World Wide Web http://www.slcc.edu/lr/navigator/discovery/discover.html. Developed by a consortium of information professionals for an online course for Internet instruction under a grant from the Higher Education Technology Initiative, State of Utah).

Evaluate the format. Can you clearly identify what type of information it is?
Is it a Web Home page?
Is it a Gopher?
Is it a newsgroup posting?
Is it a government report?
Is it an advertisement? (Janet Hovorka and Keith Slade, "Evaluating and Citing Sources: Internet Truth or Fiction?" , http://www.slcc.edu/lr/library/intwork/intwork.html Salt Lake Community College.)

3) Validation:

The discerning student should also continue the critical information of sources by examining the criteria of the author. For example:
What can you learn about the author's reputation?
Does the information presented appear accurate and objective?
What is the political, cultural, religious, or disciplinary perspective of the author? For example, I have had several students use data gleaned from the Reagan Home Page. Students need to note the perspective of the source before accepting whole-heartedly the statistical representations of any author. For example, is the information from a primary or secondary source? Issues of timeliness and currency should be explored.
When was the information source created?
Is the information substantiated by other sources?
In many cases, cost becomes an issue in evaluating research materials. Researchers should ask themselves:
Is the information free or is there a fee?
If there is a charge, why?
What type of source is requesting a fee? Is it an advertisement, or a commercial venture trying to sell

mostly for profit, or is it an academic organization charging a fee merely to defray publishing costs? (C. Hansen, Ibid.)

4) Advantages of Valid Internet Research:

Bruce Worthen suggests that Internet research presents a quick and easy way to verify sources listed by students in their papers. Unlike the "old" days, when professors had to sit in the library (or send their research assistants) to spot check sources in periodicals and books, they can now sit at the comfort of their office computer and check the addresses of cited Internet works in a matter of minutes. Indeed, even Hope Tillman, Director of Libraries at Babson College, admitted:

"Sharyn Ladner and I wrote our first book surveying the Internet use of special librarians in 1991 and 1992...and we noted that ..."the Internet allows all types of publishing in the broadest sense--much of the infoormation contained in the Internet resident discussion groups is transitory--and this network of networks will continue to expand exponentially so that bibliographic control will continue to be out of reach. " (Sharyn J. Ladner and Hope N. Tillman, *Internet and Special Librarians: Use, Training, and the Future*. Washington, D.C. : Special Libraries Association, 1993, p. 58)

Indeed, what a difference a couple of years makes. The above authors admit that their crystal ball was not very good. Actually, there is the potential for a whole lot more bibliographic control today; and at the same time there is increasing complexity. Hence, more reason for information professionals' dedication to developing their skills for "search" tool development for whatever the Internet is going to become. (Hope Tillman, "Finding Quality on the Internet or a Needle in a Haystack?" in World Wide Web, http://www.tiac.net/users/hope/findqual.html INTERNET.)

Some of the search engines have developed into dependable vehicles for verification and evaluation of sources. Consider the following projects:

W3 Virtual Libraries Project

The W3 Virtual Libraries initial approach to subject guides to the Internet purported to be a scholarly one. They sought subject experts to develop annotated lists of sites in their fields, both broadly and narrowly. The problem has become the uneven quality of the guides and even the different approaches which grew out of the creativity of their developers. While there are clues on the pages, some have not been maintained and represent an initial or periodic effort rather than an ongoing one. Others are very up-to-date and complete. As the web has exploded, keeping up with these subject guides has become much more complex and difficult.

Clearinghouse Project

This project is lead by Louis Rosenfeld, a Ph.D. candidate at the University o Michigan library school. According to information on the Clearinghouse web site, he plans to rate each of the guides according to four criteria:

1. Level of resource description: descriptive information providing users with an objective sense of what an Internet resource covers
2. Level of resource evaluation: evaluative information providing users with a subjective sense of the quality of an Internet resource

3. Organizational schemes, or how the guide is organized (by subject, format, audience, or other)
4. Level of meta-information, or information about other information. For instance, information about the authors, their professional or institutional affiliations and their knowledge or experience with the subject; how the guide was researched and constructed; and the mission of the guide.

Yahoo

Originally , Yahoo was started as a project by its two co-authors who wanted to share their Web bookmarks. Although they started as graduate students at Stanford, they have since left there and reside at Netscape, where they have a staff to help them. At last glimpse, they wer advertizing for a cataloging librarian. They are soliciting URLs, categorizing the, and adding them to their database. They do not guarentee quality. However, one good feature of Yahoo is their technique of automatically polling sites to see if they are "up" or available.

In this world of meta information, or information about information, perhaps the next service to come along will be a group that provide "evaluations" of Internet groups that "evaluate" Internet resources. Are these resources that provide evaluations truly unbiased, or are they subjective in their analysis? Examine, for instance:

Point Communications

This is an independent company in New York with a staff of 10-25 reviewers. They use Lycos search engine for the point Search function. They claim no relationsip between their advertisng and their reviews of what they term "the largerst and best collection of entertaing review of the Web on the Web". Recently sold to Lycos, Point's staff claims to "...surf the Web daily looking for the best, smartest, and most entertaining sites around. If we review a page it means we think is is among the best 5% of all Web sites in content, presentation, and/or experience. Point makes no distinction between commercial, private, or student pages. Excellence is our only criterion."

The McKinley

The McKinley Internet Directory is an online directory of described, rated, and reviewed Internet resources and other key facts instantly accessible to users as they scan the result of their search in the McKinley. It uses the PLS search engine. Reviews are performed by a team of highly skilled international publishers, technologists and information specialists. According to the information on its web menu, "The McKinley currently contains over 20,000 evaluated, reviewed and rated sites, of which approximately 35 percent are international in origin... Rating System: The star rating that appears near the top of each review is an average of the ratings from each of four categories (4 stars is the maximum rating). HTTP, Gopher, FTP and Telnet rating measure: completeness of content presented in the resource, organization of the resource, up-to-date-ness of the information presented, and ease of access to the resource. Because of their differing functions, the ratings assigned to newsgroups, mailing lists and listservs reflect a slightly different system than the ratings for the other sites mentioned above..."

Currently, the McKinley is free. It has been licensed by the internet provider Netcom and also by IBM for use in its infomarket service.

Gale Guide

This web site is an example of a publisher offering updated information online as a supplement to their print publication. It also has descriptive information for 145 specialized home pages. (Hope Tillman, Ibid.)

Structure and Search Technique

Ian R. Winship, in his research at the Information Services Department, University of Northumbria at Newcastle, UK had postulated that, prior to his investigation concerning Internet research evaluation, retrieval performance would be of primary importance. However, he found that *record structure* and *technique* rose in greater significance as he toured the Web landscape. For example, to get some indication of the practical value of the different search engines, test searches on three subjects were carried out. The main topics he selected were:

1. The ebola virus: a specific subject for which there was not expected to be a huge amount of information
2. Tourism in Alberta, Canada: a less academic subject
3. Jacques Chirac: a non-U.S. topic

Table One gives the number of items found.

TABLE ONE

	Worm	WebCrawler	Lycos	Harvest	Galaxy	Yahoo
ebola	27	124	295	17	11	7
Alberta	0?	42	42	4	6	0
Chirac	0?	7	27	2	0	0

The zeros for the Worm are queried because that system tells you that if you get a zero response it may be because the computer is too busy to process your request and not because there is nothing relevant.

It may be worthy to note that when the ebola search was repeated on Lycos two weeks later, there were 504 items!

Winship warns that an immediate response must not be to assume that only Lycos is of real use. Analysis of results shows excessive duplication of sources with many of them at best of marginal interest. There may seem to be no more than 10 major collections of information on ebola, but hundreds of references to these from other related or personal homepages. Indeed the services with scoring all had only 6 or 7 items in the top half of the scoring range. Consequently the appearance

of the word 'ebola' in a document title or as part of the URL is more likely to indicate a precise hit than when it is in the text of a page. Therefore, tools that search only for these will give good results. Services like Galaxy and Yahoo have a more structured collection of sources that should, in information retrieval tems, give lower recall, but higher precision. These services would often advise checking their classified groupings first. When there is no source specifically on a topic, as in the Chirac example, then these are less helpful. (Ian R. Winship, "World Wide Web Searching Tools - An Evaluation." VINE (99) 1995, 49-54; South Bank University, Library Information Technology Centre, London.)

One may agree with Winship in his contention that it may be more fruitful to use browsable collections like the BUBL Subject Tree, especiallly if they also include gopher material, which is becoming too easily overlooked in the Web dominated world. We should also remember that these servicees are not intended for information professionals per se and despite their deficiencies in documentation and structure, they are very popular. This leads us to the questions:

Should teachers and information professionals get involved in the design of search tools to make them more effective and usable?

Can these tools be incorporated into the mainstream of online searching as we know it?

Is it the responsibility of teachers using Internet to provide clear instruction on evaluation and validation of sources before assigning research projects to students?

Do students and teachers have an ethical responsibility to validate all sources used in scholastic research, so as to build a new culture of tradition for Internet use? Or, is this just more unnecessary procedure? Won't the information just take care of itself, or stand on its own merits?

CONCLUSION

As you can see, the Internet has exploded into a place where many of the traditional rules of research and scholarship have changed significantly. In the case of homepage proliferation, gone are the publishing house rules of jury and peer review. Slipping fast are the memories of old fashioned research using the Reader's Guide to Periodical Literature in text form, and the microfiche parties in the dungeons of libraries, standing in line with pockets of quarters behind other totally irritated students and researchers, waiting a turn to obtain a poor copy of some possibly outdated piece of information. And in many places, the card catalog has been replaced by the computer, with the distant possibility of some Luddite * rebels entering the libraries one day, smashing the monitors in defiance of this information takeover.

Yes, the rules have changed dramatically, but have the teachers and information professionals moved at the same rate of speed? Should there be some task force, some self-appointed cadre of beings, who will arise as the masters of information validation?

Imagine, for a moment, a world in which all teachers and information professionals turned their backs to these questions, and paid no mind to the ethical considerations of an untended information explosion. Perhaps some utopian scenario may arise, in which information continues to float in cyberspace, available to many, yet discernable by only a few, and harmless to all. A more realistic situation, however, might find our students subject to more massive manipulation by the media and the information industry.

* Inspired by Steven Ruffus, Professor of English and co-author of English 101 On-Line, a writing course offered on the World Wide Web and funded by the Higher Education Technology Initiative, State of Utah; http://www.slcc.edu/b10/eng101.html

REFERENCES

December, John. "Internet Tools Summary" in World Wide Web, 1995. Available from http://www.rpi.edu/Internet/Guides/decemj/itools/top.html INTERNET.

Dillon, Martin et. al. "Assessing Information on the Internet; Toward Providing Library Services for Computer Mediated Communication" in World Wide Web. Available from http:www/oclc.org:5047/oclc/research/publications/aii/table.html INTERNET.

Hanson, C. "Internet Navigator - Resource Discovery" in World Wide Web. Available from http://www.slcc.edu/lr/navigator/discovery/discover.html. Developed by a consortium of information professionals led by Ms. Hanson for an online course for internet instruction under a grant from the Higher Education Technology Initiative, State of Utah.

Hovorka, Janet and Slade, Keith. "Evaluating and Citing Sources: Truth or Fiction?" In World Wide Web, 1996. Available from http://www.slcc.edu/lr/library/intwork/intwork.htm

Ladner, Sharon and Tillman, Hope. *Internet and Special librarians: Use, Training, and the Future*. Washington, D.C.: Special Libraries Association, 1993, p.58).

Large, J.A. "Evaluating Online and CD-ROM Sources." *Journal of Librarianship* 21 (2) April 1989, 87-108.

Tillman, Hope. "Finding Quality on the Internet or a Needle in a Haystack?" in World Wide Web (Massachusetts, September 6, 1995). Available from http://www.tiac.net/users/hope/findqual.html INTERNET.

Winship, Ian R. "World Wide Web Searching Tools - An Evaluation." *VINE* (99) 1995, 49-54. Library Information Technology Centre, South Bank University, London. Also available from http://www.bubl.bath.ac.uk/BUBL/IWinship.html